# Improved Converses and Gap-Results for Coded Caching

Chien-Yi Wang, Shirin Saeedi Bidokhti, and Michèle Wigger

**Abstract**

Improved lower bounds on the average and the worst-case rate-memory tradeoffs for the Maddah-Ali&Niesen coded caching scenario are presented. For any number of users and files and for arbitrary cache sizes, the multiplicative gap between the exact rate-memory tradeoff and the new lower bound is less than 2.315 in the worst-case scenario and less than 2.507 in the average-case scenario.

## I. Introduction

We consider the canonical coded caching scenario by Maddah-Ali and Niesen [1] with a single transmitter and $K$ receivers, where each receiver is equipped with a cache memory of equal size (see in Figure 1). Communication takes place in two phases. In a first *caching phase* the transmitter stores contents (arbitrary functions of files) at the receivers' cache memories. In a subsequent *delivery phase* it conveys one of the files stored in its library to each of the receivers. The main challenge in this configuration is that during the caching phase it is not known which receiver demands which specific file from the library. The cache contents thus need to be designed so as to be useful for many possible demands.

Traditional caching systems store the same most popular files in each and every cache memory. This allows the receivers to retrieve these files locally without burdening the common communication link from the transmitter to the receivers. Recently [1], it has been shown that much larger gains, so called *global caching gains*, are possible if the various receivers store different parts of the files in their cache memories. In this case, the transmitter can simultaneously serve multiple receivers during the delivery phase by sending coded data, and thus significantly reduced the required communication rate on the common link connecting the transmitter to the receivers.

The main quantity of interest in this work is the *rate-memory tradeoff*—i.e., the minimum required delivery rate, as a function of the available cache memories, so that all receivers reliably recover their demanded files. Several recent works [1]–[17] have presented upper and lower bounds on the rate-memory tradeoff. More specifically, the works in [1]–[12], [14] investigated the worst-case rate-memory tradeoff where the communication rate *cannot* depend on the specific demand. (The works in [4], [5], [14] determined the exact rate-memory tradeoff if the caching contents are restricted to parts of messages. This caching strategy is however known to be suboptimal in general.) The currently best lower and upper bounds for this worst-case rate-memory tradeoff have been shown to match up to a multiplicative gap of $4$ [10].

The works in [13], [14] investigated the average rate-memory tradeoff where the communication rate *can* depend on the specific demand, and the rate of interest is the average required rate over all possible demands. The currently best lower and upper bounds for this average rate-memory tradeoff have been shown to match up to a multiplicative gap of $4.7$ [13].

In this paper we provide new lower bounds on the worst-case and the average rate-memory tradeoffs. The new lower bounds match the exact worst-case and average rate-memory tradeoffs up to multiplicative gaps of $2.315$ and $2.507$, respectively. More precisely, these gaps are with respect to the upper bounds on the rate-memory tradeoffs under *decentralized* caching in [14]. An upper bound on the rate-memory tradeoff under decentralized caching is also an upper bound on the rate-memory tradeoff under *centralized* caching considered here, because in the decentralized caching the cache content at a given receiver has to be chosen according to a specific distribution, whereas in centralized caching any content can be cached that satisfies the cache memory constraints.

C.-Y. Wang and M. Wigger are with LTCI, Telecom ParisTech, Université Paris-Saclay, 75013 Paris, michele.wigger@telecom-paristech.fr. S. Saeedi Bidokhti is with the Department of Electrical Engineering at Stanford University, saeedi@stanford.edu.

Parts of the material in this paper were submitted to the *2017 IEEE International Symposium on Information Theory*, Aachen, Germany.

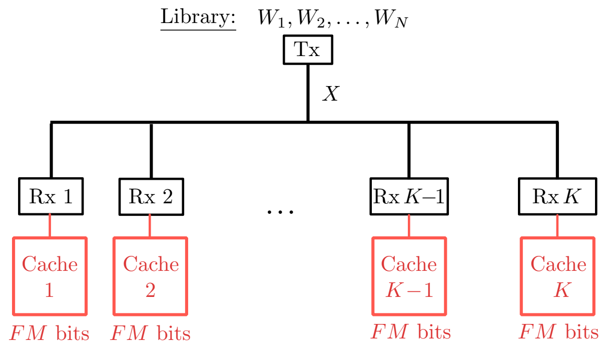


Fig. 1: Coded caching scenario with $K$ receivers having equal cache size $FM$ bits.

Many other variations of the caching problem have been studied. For example, the *online caching problem* [18]; caching with non-uniform demands [19]–[21]; caching of correlated files [22]–[24] where Wyner's and Gac-Körner's common information play an important role [22]; caching in *noisy* broadcast channels [25]–[27], [29]–[43] where perfomance can be improved through joint cache-channel coding [25], [26], [32], state-information [33], [34], [34], [37], [39], or massive MIMO [36]; caching in Gaussian interference networks [44]–[46]; caching in hierarchical networks [54], [56], [57]; and caching in cellular networks [46]–[53] where it allows to cancel inter-cell interference [47].

## II. Detailed Problem Setup

The transmitter has a library of $N$ independent messages $W_1, \ldots, W_N$, where each $W_d$ is uniformly distributed over $\{1, \ldots, 2^F\}$ for $F$ a positive integer. Each of the $K$ receivers is provided with a cache memory of size $F\mathsf{M}$ bits, see Figure 1. Suppose that each receiver will demand exactly one message from the library. We denote the demand of receiver $k$ by

$$d_k \in \mathcal{N} := \{1, \ldots, N\}.$$

and thus the message demanded by receiver $k$ is $W_{d_k}$. Let

$$\mathbf{d} := (d_1, \ldots, d_K)$$

denote the receivers' demand vector. The communications process takes place in two phases: a caching phase and a delivery phase.

During a period of low network-congestion and before the receivers' demand vector $\mathbf{d}$ is known, the transmitter sends an individual cache message $\mathbb{V}_k \in \left\{1, \ldots, \left\lfloor 2^{F\mathsf{M}} \right\rfloor\right\}$, to each of the $K$ receivers. Since $\mathbf{d}$ is unknown at this time, the cache messages will be functions of the entire library. For every $k \in \{1, \ldots, K\}$:[1]

$$\begin{aligned} \mathbb{V}_k &:= g_k(W_1, \ldots, W_N), \\ g_k : &\quad \{1, \ldots, 2^F\}^N \to \{1, \ldots, \lfloor 2^{F\mathsf{M}} \rfloor\}. \end{aligned}$$

In the delivery phase, the transmitter is given the receivers' demands $\mathbf{d} = (d_1, \ldots, d_K)$, and it generates the delivery-symbol $X$ sent over the common noise-free bit-pipe as:

$$\begin{aligned} X &:= f_{\mathbf{d}}(W_1, \ldots, W_N), \\ f_{\mathbf{d}} : &\quad \{1, \ldots, 2^F\}^N \to \mathcal{X}, \end{aligned}$$

where $\mathcal{X}$ is the delivery alphabet that we will specify shortly.

We also assume that $\mathbf{d}$ is known to all receivers (e.g., $\mathbf{d}$ can be communicated to the receivers with zero transmission rate). Each receiver $k \in \{1, \ldots, K\}$ perfectly observes the delivery-symbol $X$, and can thus recover its desired message as

$$\hat{W}_k := \varphi_{k,\mathbf{d}}(X, \mathbb{V}_k),$$

[1]Allowing for randomized caching functions does not change the results of this paper.

$$\varphi_{k,\mathbf{d}}: \quad \mathcal{X} \times \{1,\dots,\lfloor 2^{F\mathsf{M}} \rfloor\} \to \{1,\dots,2^F\}.$$

We are left with specifying the delivery alphabet $\mathcal{X}$. We distinguish two scenarios:

- In the *worst-case scenario*, the delivery alphabet $\mathcal{X}$ does not depend on the demand vector $\mathbf{d}$.
  In this scenario, a rate-memory pair $(R,\ \mathsf{M})$ is *achievable* if for every $\epsilon > 0$ and sufficiently large message lengths $F$, there exist caching, encoding, and decoding functions with delivery alphabet
  $$\mathcal{X} = \{1,\dots,\lfloor 2^{F(R+\epsilon)} \rfloor\}, \tag{1}$$
  so that for each demand vector $\mathbf{d} \in \mathcal{N}^K$, each receiver $k$ can perfectly reconstruct its desired message:
  $$\hat{W}_k = W_{d_k}. \tag{2}$$
- In the *average-case scenario*, the delivery alphabet $\mathcal{X}$ depends on the demand vector $\mathbf{d}$.
  In this scenario, a rate-memory pair $(R,\ \mathsf{M})$ is *achievable* if for each demand vector $\mathbf{d} \in \mathcal{N}^K$, each $\epsilon > 0$, and sufficiently large message lengths $F$, there exist caching, encoding, and decoding functions with delivery alphabet
  $$\mathcal{X}_{\mathbf{d}} = \{1,\dots,\lfloor 2^{FR_{\mathbf{d}}} \rfloor\}, \tag{3}$$
  so that each Receiver $k \in \{1,\dots,K\}$ can perfectly reconstruct its desired message, (2), and
  $$\frac{1}{N^K} \sum_{\mathbf{d} \in \mathcal{N}^K} R_{\mathbf{d}} \leq R + \epsilon. \tag{4}$$

The main focus of this paper is on the *rate-memory tradeoffs* of the worst-case and the average-case scenarios:

*Definition 1:* Given the cache memory size $\mathsf{M}$, we define the *rate-memory tradeoffs* $R^\star_{\text{worst}}(\mathsf{M})$ and $R^\star_{\text{avg}}(\mathsf{M})$ as the infimum of all rates $R$ such that the rate-memory pair $(R,\ \mathsf{M})$ is achievable for the worst-case and the average-case scenarios, respectively.

The problem is interesting for
$$\mathsf{M} < N. \tag{5}$$
For $\mathsf{M} \geq N$ each receiver can store all the library in its cache memory and there is no need to transmit anything during the delivery phase. We henceforth assume that (5) holds.

## III. Main Results

Define $\bar{N} := \min\{K, N\}$ and $\bar{\mathcal{N}} := \{1, 2, \dots, \bar{N}\}$.

### A. Worst-Case Scenario

Our first result is a lower bound on the rate-memory tradeoff in the worst-case scenario. It is proved in Section IV-B. Alternatively, it can also be extracted from the converse result for general degraded broadcast channels in [28].

*Theorem 1:* For all $\mathsf{M} \in [0, N)$,
$$R^\star_{\text{worst}}(\mathsf{M}) \geq R^{\text{low}}_{\text{worst}}(\mathsf{M}), \tag{6}$$
where
$$R^{\text{low}}_{\text{worst}}(\mathsf{M}) := \max\Bigg\{ \max_{\ell \in \bar{\mathcal{N}}} \left[ \ell - \mathsf{M}\frac{\ell^2}{N} \right], \max_{\ell \in \bar{\mathcal{N}}} \left[ \ell - \mathsf{M} \sum_{k=1}^{\ell} \frac{k}{N-k+1} \right] \Bigg\}.$$

Figure 2 compares this new lower bound on $R^\star_{\text{worst}}(\mathsf{M})$ with the existing lower bounds in [1] and [6]. The figure also shows upper bounds from [14]. The red solid upper bound is for centralized caching, as considered in this paper. The green dashed upper bound is for decentralized caching. For simplicity, the latter upper bound is used to derive the gap results in the following Theorem 2.

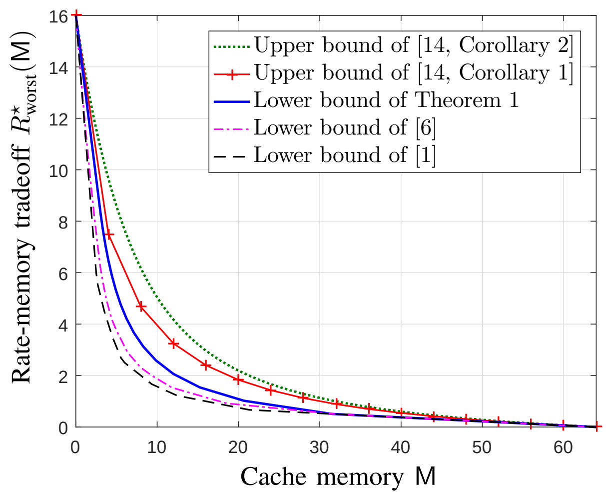


Fig. 2: Upper and lower bounds on $R^{\star}_{\text{worst}}(\mathsf{M})$ for $K = 16$ and $N = 64$.

*Theorem 2:* Irrespective of the number of users $K$, the library size $N$, and the memory size $\mathsf{M} \in [0, N)$:

$$\frac{R^{\star}_{\text{worst}}(\mathsf{M})}{R^{\text{low}}_{\text{worst}}(\mathsf{M})} \leq \max_{\ell \in \mathbb{Z}^+} \max_{a \in (0,1)} \phi(a, \ell),$$

where

$$\phi(a,\ell) := \frac{\frac{a(\ell+1)}{(1-a)\ell}\left(1 - \left(\frac{\ell+a}{\ell+1}\right)^{\ell/a}\right)}{1-(1-a)\sum_{j=0}^{\ell-1}\frac{1}{\ell-aj}} \leq 2.315$$

*Proof:* See Section V. ■

*Remark 3:* For any $\ell \in \mathbb{Z}^+$, the function $a \mapsto \phi(a, \ell)$ is continuous and bounded over $(0, 1)$, see also Figure 3. Numerical evaluation[2] shows that for $\ell \in \{1, \dots, 10^4\}$:

$$\max_{a \in (0,1)} \phi(a, \ell) \leq 2.315.$$

Moreover,

$$\max_{\ell \in \mathbb{Z}^+ :\, \ell > 10^4} \max_{a \in (0,1)} \phi(a, \ell) \leq \max_{b \in (0, 10^{-4})} \max_{a \in (0,1)} \psi(a, b), \tag{7}$$

where

$$\psi(a,b) := \frac{\frac{a(1+b)}{1-a}\left(1 - \left(\frac{1+ab}{1+b}\right)^{\frac{1}{ab}}\right)}{1 - \frac{(1-a)b}{1-a+ab} + \frac{1-a}{a}\ln(1-a+ab)}. \tag{8}$$

The function $(a, b) \mapsto \psi(a, b)$ is continuous and bounded over $(0, 1) \times (0, 10^{-4})$, see also Figure 4. Numerical evaluation shows that

$$\max_{b \in (0, 10^{-4})} \max_{a \in (0,1)} \psi(a, b) \leq 2.315. \tag{9}$$

*Proof:* Inequality (7) is proved in Section V-B. ■

[2]All numerical evaluations in this paper are performed by applying the MATLAB function `fmincon` with the sequential quadratic programming (SQP) method.

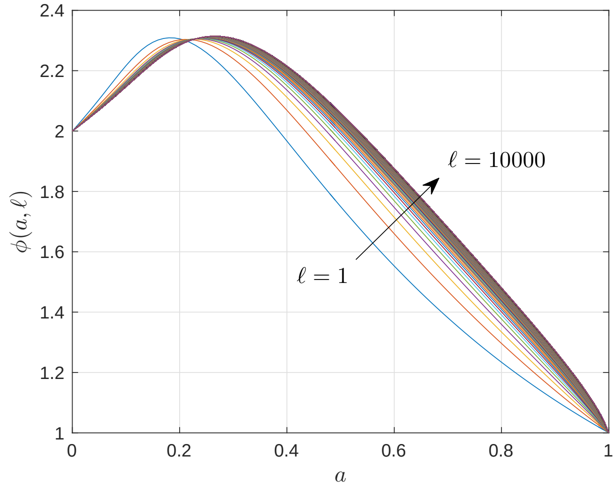


Fig. 3: The functions $\phi(a,\ell)$ for $a \in (0,1)$ and $\ell = 1, \ldots, 10^4$.

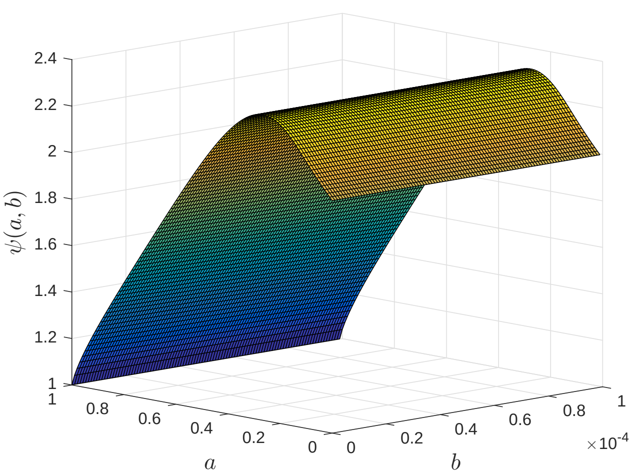


Fig. 4: The function $\psi(a,b)$ over $(a,b) \in (0,1) \times (0,10^{-4})$.

### B. Average-Case Scenario

*Theorem 4:* For all $\mathsf{M} \in [0,N)$,

$$R^{\star}_{\text{avg}}(\mathsf{M}) \geq R^{\text{low}}_{\text{avg}}(\mathsf{M}),$$

where

$$\begin{aligned} &R^{\text{low}}_{\text{avg}}(\mathsf{M}) \\ &:= \max\left\{ \max_{\ell \in \{1,\ldots,K\}} \left[ \left(1 - \left(1 - \frac{1}{N}\right)^{\ell}\right)(N - \ell\mathsf{M}) \right], \right. \\ &\qquad \left. \max_{\ell \in \{1,\ldots,K\}} \left[ \left(1 - \left(1 - \frac{1}{N}\right)^{\ell}\right) N - \frac{\ell(\ell+1)}{2N}\mathsf{M} \right] \right\}. \end{aligned}$$

*Proof:* See Section IV-C. ■

Figure 5 compares this new lower bound on $R^{\star}_{\text{avg}}(\mathsf{M})$ with the existing lower bounds in [13] and the upper bounds in [14]. The red solid upper bound is for centralized caching, as considered in this paper. The green dashed upper bound is for decentralized caching and also from [14].

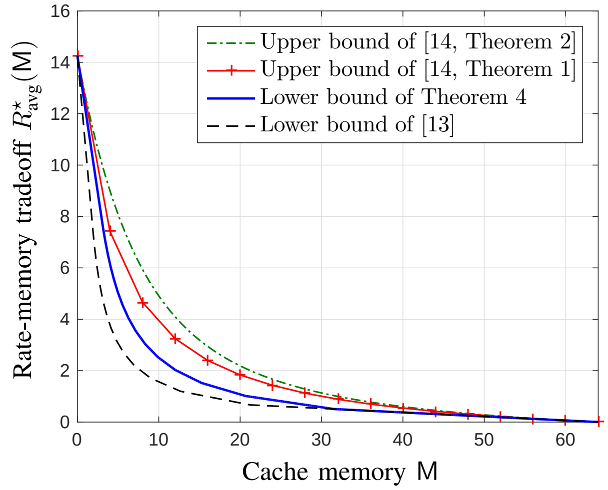


Fig. 5: Upper and lower bounds on $R^\star_{\text{avg}}(\mathsf{M})$ for $K = 16$ and $N = 64$.

As the following theorem and remark show, the multiplicative gap between the lower bound of Theorem 4 and $R^\star_{\text{avg}}(\mathsf{M})$ is at most 2.507.

*Theorem 5:* Irrespective of the number of users $K$, the library size $N$, and the memory size $\mathsf{M} \in [0, N)$:

$$\frac{R^\star_{\text{avg}}(\mathsf{M})}{R^{\text{low}}_{\text{avg}}(\mathsf{M})} \leq \max_{u\in(0,1]} \max_{v\in(0,1/2]} \eta(u,v),$$

where

$$\eta(u,v) := \frac{\left(u+v-v\left(1-v\right)^{\frac{u}{v}}\right)\left(1-\left(1-\frac{v}{u+v}\left(1-v\right)^{\frac{u}{v}}\right)^{\frac{1}{v}}\right)}{\left(1-v\right)^{\frac{u}{v}}\left(1-\left(1+\frac{u}{2}\right)\left(1-v\right)^{\frac{u}{v}}\right)}.$$

*Proof:* See Section VI. ■

*Remark 6:* The function $\eta$ is continuous and bounded over $(0,1] \times (0,1/2]$. Numerical evaluation shows that

$$\max_{u\in(0,1]} \max_{v\in(0,1/2]} \eta(u,v) \leq 2.507. \tag{10}$$

## IV. Proof of Theorems 1 and 4

### A. Auxiliary Lemmas

The following two lemmas will be used in the proofs of Theorems 1 and 4.

The next lemma is stated for the average-case scenario. It applies readily also to the worst-case scenario if rate $R_{\mathbf{d}}$ is replaced by $R$.

*Lemma 7:* Fix a number $\ell \in \bar{\mathcal{N}}$ and a demand vector $\mathbf{d} \in \mathcal{N}^K$ whose first $\ell$ entries are $d_1, \ldots, d_\ell$. Fix also a small $\epsilon > 0$ and assume a sufficiently large $F$ with caching, encoding, and decoding functions so that (2) holds for all $k \in \{1, \ldots, K\}$. Then,

$$R_{\mathbf{d}} + \epsilon \geq \kappa_{\mathbf{d}}(\ell) - \frac{1}{F}\sum_{k=1}^{\ell} I(W_{d_k}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{d_1}, \ldots, W_{d_{k-1}}), \tag{11}$$

where $\kappa_{\mathbf{d}}(\ell)$ denotes the number of distinct demands for receivers $1, \ldots, \ell$:

$$\kappa_{\mathbf{d}}(\ell) := |\{d_1, \ldots, d_\ell\}|. \tag{12}$$

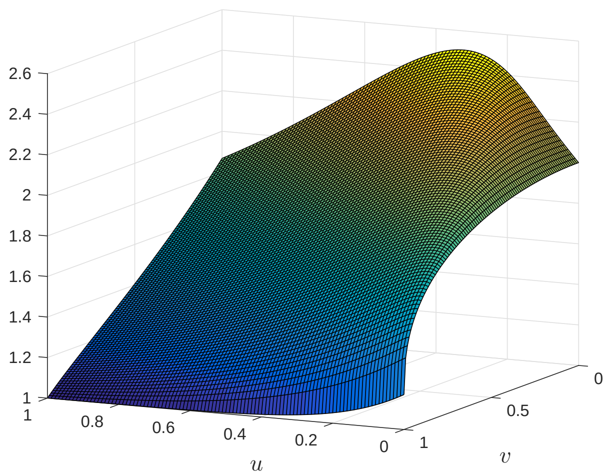


Fig. 6: The function $\eta(u,v)$ over $(u,v) \in (0,1) \times (0,1)$.

*Proof:* For any $k \in \{1, \ldots, \ell\}$:

$$
\begin{aligned}
&I(X; W_{d_k} | \mathbb{V}_1, \ldots, \mathbb{V}_k, W_{d_1}, \ldots, W_{d_{k-1}}) \\
&\quad \overset{(a)}{=} H(W_{d_k} | \mathbb{V}_1, \ldots, \mathbb{V}_k, W_{d_1}, \ldots, W_{d_{k-1}}) \\
&\quad = H(W_{d_k} | W_{d_1}, \ldots, W_{d_{k-1}}) \\
&\qquad - I(W_{d_k}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{d_1}, \ldots, W_{d_{k-1}}) \\
&\quad \overset{(b)}{=} F \cdot \mathbb{1}\big\{d_k \notin \{d_1, \ldots, d_{k-1}\}\big\} \\
&\qquad - I(W_{d_k}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{d_1}, \ldots, W_{d_{k-1}}),
\end{aligned} \tag{13}
$$

where (a) holds because (2) implies that $H(W_{d_k} | X, \mathbb{V}_1, \ldots, \mathbb{V}_k, W_{d_1}, \ldots, W_{d_{k-1}}) = 0$; and (b) holds by the independence of the messages and because $H(W_d) = F$ for any $d \in \mathcal{N}$.

On the other hand,

$$
\begin{aligned}
&\sum_{k=1}^{\ell} I(X; W_{d_k} | \mathbb{V}_1, \ldots, \mathbb{V}_k, W_{d_1}, \ldots, W_{d_{k-1}}) \\
&\quad \leq \sum_{k=1}^{\ell} I(X; W_{d_k}, \mathbb{V}_k | \mathbb{V}_1, \ldots, \mathbb{V}_{k-1}, W_{d_1}, \ldots, W_{d_{k-1}}) \\
&\quad = I(X; W_{d_1}, \ldots, W_{d_\ell}, \mathbb{V}_1, \ldots, \mathbb{V}_\ell) \\
&\quad \leq H(X) \leq F(R_{\mathbf{d}} + \epsilon).
\end{aligned} \tag{14}
$$

Combining (13) and (14) establishes the lemma. ■

*Lemma 8:* Let $L$ be a positive integer, $A_1, \ldots, A_L$ be an independent random $L$-tuple, and $\mathbb{V}$ be a random variable arbitrarily correlated with $A_1, \ldots, A_L$. For any subset $\mathcal{S} \subseteq \{1, \ldots, L\}$, denote by $A_{\mathcal{S}}$ the subset $\{A_s,\ s \in \mathcal{S}\}$. Then, for all $l \in \{1, \ldots, L\}$,

$$
\frac{1}{\binom{L}{l}} \sum_{\substack{\mathcal{S} \subseteq \{1, \ldots, L\}: \\ |\mathcal{S}| = l}} I(A_{\mathcal{S}}; \mathbb{V}) \leq \frac{l}{L} I(A_1, \ldots, A_L; \mathbb{V}). \tag{15}
$$

*Proof:* Consider any $l \in \{1, \ldots, L\}$. We have

$$
\frac{1}{\binom{L}{l}} \sum_{\substack{\mathcal{S} \subseteq \{1, \ldots, L\}: \\ |\mathcal{S}| = l}} I(A_{\mathcal{S}}; \mathbb{V})
$$

$$\stackrel{(a)}{=} \frac{1}{\binom{L}{l}} \sum_{\substack{\mathcal{S}\subseteq\{1,\ldots,L\}:\\ |\mathcal{S}|=l}} \sum_{j\in\mathcal{S}} H(A_j)$$

$$-\frac{1}{\binom{L}{l}} \sum_{\substack{\mathcal{S}\subseteq\{1,\ldots,L\}:\\ |\mathcal{S}|=l}} H(A_{\mathcal{S}}|\mathbb{V})$$

$$= \frac{\binom{L-1}{l-1}}{\binom{L}{l}} \sum_{j=1}^{L} H(A_j) - \frac{1}{\binom{L}{l}} \sum_{\substack{\mathcal{S}\subseteq\{1,\ldots,L\}:\\ |\mathcal{S}|=l}} H(A_{\mathcal{S}}|\mathbb{V})$$

$$\stackrel{(b)}{\leq} \frac{l}{L}\sum_{j=1}^{L} H(A_j) - \frac{l}{L} H(A_1,\ldots,A_L|\mathbb{V})$$

$$\stackrel{(c)}{=} \frac{l}{L} I(A_1,\ldots,A_L;\mathbb{V}),$$

where $(a)$ and $(c)$ follow since $A_1,\ldots,A_L$ are independent and $(b)$ follows from the generalized Han Inequality (see [60, Theorem 17.6.1]). ■

### *B. Proof of Theorem 1*

Fix $\ell \in \bar{\mathcal{N}}$ and restrict attention to receivers $1,\ldots,\ell$ and their cache memories. Let $\mathcal{Q}_\ell^{\text{dist}}$ be the set of all *ordered* $\ell$-dimensional demand vectors $(d_1,\ldots,d_\ell)$ with all distinct entries. So,

$$|\mathcal{Q}_\ell^{\text{dist}}| = \binom{N}{\ell}\, \ell!. \tag{16}$$

Notice that for $\mathbf{d} \in \mathcal{Q}_\ell^{\text{dist}}$, we have $\kappa_{\mathbf{d}}(\ell) = \ell$, and averaging Inequality (11)[3] over all demand vectors $\mathbf{d} \in \mathcal{Q}_\ell^{\text{dist}}$ yields the following inequality:

$$R + \epsilon \geq \ell - \sum_{k=1}^{\ell} \alpha_k, \tag{17}$$

where

$$\alpha_1 := \frac{1}{\binom{N}{\ell}\ell!} \sum_{\mathbf{d}\in\mathcal{Q}_\ell^{\text{dist}}} \frac{1}{F} I(W_{d_1};\mathbb{V}_1), \tag{18a}$$

and for $k = 2,\ldots,\ell$:

$$\alpha_k := \frac{1}{\binom{N}{\ell}\ell!} \sum_{\mathbf{d}\in\mathcal{Q}_\ell^{\text{dist}}} \frac{1}{F} I(W_{d_k};\mathbb{V}_1,\ldots,\mathbb{V}_k|W_{d_1},\ldots,W_{d_{k-1}}). \tag{18b}$$

The following lemma and letting $\epsilon \to 0$ and thus $F \to \infty$, concludes the proof.

*Lemma 9:* Parameters $\alpha_1,\ldots,\alpha_\ell$ satisfy

$$\sum_{k=1}^{\ell} \alpha_k \leq \min\left\{ \frac{\ell^2}{N}\mathsf{M},\ \sum_{k=1}^{\ell} \frac{k\mathsf{M}}{N-k+1} \right\}. \tag{19}$$

*Proof:* We first prove that for each $k \in \{1,\ldots,\ell\}$:

$$\alpha_k \leq \frac{k\mathsf{M}}{N-k+1}, \tag{20}$$

[3] In (11) $R_{\mathbf{d}}$ needs to be replaced by $R$ because here we consider a worst-case scenario.

which establishes the upper bound

$$\sum_{k=1}^{\ell} \alpha_k \le \sum_{k=1}^{\ell} \frac{k\mathsf{M}}{N-k+1}. \tag{21}$$

For each partial demand vector $\tilde{\mathbf{d}} = (d_1, \ldots, d_{k-1})$, let $W_{\tilde{\mathbf{d}}} := \{W_{d_1}, \ldots, W_{d_{k-1}}\}$. We have:

$$\begin{aligned}
&F\alpha_k \\
&= \frac{1}{\ell!\binom{N}{\ell}} \sum_{\mathbf{d}\in\mathcal{Q}_\ell^{\text{dist}}} I(W_{d_k}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{d_1}, \ldots, W_{d_{k-1}}) \\
&= \frac{1}{\ell!\binom{N}{\ell}} \sum_{\tilde{\mathbf{d}}\in\mathcal{Q}_{k-1}^{\text{dist}}} \sum_{\substack{\mathbf{d}\in\mathcal{Q}_\ell^{\text{dist}}: \\ (d_1,\ldots,d_{k-1})=\tilde{\mathbf{d}}}} I(W_{d_k}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{\tilde{\mathbf{d}}}) \\
&\overset{(a)}{=} \frac{1}{\ell!\binom{N}{\ell}} \sum_{\tilde{\mathbf{d}}\in\mathcal{Q}_{k-1}^{\text{dist}}} \sum_{j\in\mathcal{N}\setminus\tilde{\mathbf{d}}} I(W_j; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{\tilde{\mathbf{d}}}) \\
&\qquad\qquad \cdot \binom{N-k}{\ell-k}(\ell-k)! \\
&= \frac{1}{k!\binom{N}{k}} \sum_{\tilde{\mathbf{d}}\in\mathcal{Q}_{k-1}^{\text{dist}}} \sum_{j\in\mathcal{N}\setminus\tilde{\mathbf{d}}} I(W_j; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{\tilde{\mathbf{d}}}) \\
&\overset{(b)}{=} \frac{1}{k!\binom{N}{k}} \sum_{\tilde{\mathbf{d}}\in\mathcal{Q}_{k-1}^{\text{dist}}} \big[H(W_1, \ldots, W_N | W_{\tilde{\mathbf{d}}}) - \\
&\qquad\qquad \sum_{j\in\mathcal{N}\setminus\tilde{\mathbf{d}}} H(W_j | \mathbb{V}_1, \ldots, \mathbb{V}_k, W_{\tilde{\mathbf{d}}})\big] \\
&\overset{(c)}{\le} \frac{1}{k!\binom{N}{k}} \sum_{\tilde{\mathbf{d}}\in\mathcal{Q}_{k-1}^{\text{dist}}} I(W_1, \ldots, W_N; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{\tilde{\mathbf{d}}}) \\
&\overset{(d)}{\le} \frac{(k-1)!\binom{N}{k-1}}{k!\binom{N}{k}} kF\mathsf{M} \\
&= \frac{kF\mathsf{M}}{N-k+1},
\end{aligned} \tag{22}$$

where $(a)$ holds because for each value of $\ell$ and $j$ there are $\binom{N-k}{\ell-k}(\ell-k)!$ ordered demand vectors $\mathbf{d} = (d_1, \ldots, d_K) \in \mathcal{Q}_\ell^{\text{dist}}$ with $(d_1, \ldots, d_{k-1}) = \tilde{\mathbf{d}}$ and with $d_k = j$; (b) holds by the independence of the messages; (c) holds because for any random tuple $(A_1, \ldots, A_L)$ it holds that $\sum_{l=1}^{L} H(A_l) \ge H(A_1, \ldots, A_L)$; and (d) holds because $I(W_1, \ldots, W_N; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{\tilde{\mathbf{d}}})$ cannot exceed $kF\mathsf{M}$. This concludes the proof of (20) and thus of (21).

We now prove

$$\sum_{k=1}^{\ell} \alpha_k \le \frac{\ell^2\mathsf{M}}{N}. \tag{23}$$

For each $\mathbf{d} \in \mathcal{Q}_\ell^{\text{dist}}$:

$$\begin{aligned}
&I(W_{d_1}; \mathbb{V}_1) + \sum_{k=2}^{\ell} I(W_{d_k}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{d_1}, W_{d_2}, \ldots, W_{d_{k-1}}) \\
&\quad \le I(W_{d_1}, W_{d_2}, \ldots, W_{d_\ell}; \mathbb{V}_1, \ldots, \mathbb{V}_\ell).
\end{aligned} \tag{24}$$

So,

$$F\binom{N}{\ell}\ell! \cdot \sum_{k=1}^{\ell} \alpha_k$$

$$
\begin{aligned}
&= \sum_{\mathbf{d}\in\mathcal{Q}_\ell^{\text{dist}}} \Bigg[ I(W_{d_1};\mathbb{V}_1) \\
&\qquad\qquad + \sum_{k=2}^{\ell} I(W_{d_k};\mathbb{V}_1,\ldots,\mathbb{V}_k|W_{d_1},W_{d_2},\ldots,W_{d_{k-1}}) \Bigg] \\
&\le \sum_{\mathbf{d}\in\mathcal{Q}_\ell^{\text{dist}}} I(W_{d_1},W_{d_2},\ldots,W_{d_\ell};\mathbb{V}_1\ldots,\mathbb{V}_\ell) \\
&= \ell! \sum_{\substack{\mathbf{d}\in\mathcal{Q}_\ell^{\text{dist}}:\\ d_1<d_2\cdots<d_\ell}} I(W_{d_1},W_{d_2},\ldots,W_{d_\ell};\mathbb{V}_1\ldots,\mathbb{V}_\ell) \\
&\overset{(a)}{\le} \ell!\binom{N}{\ell}\frac{\ell}{N} I(W_1,\ldots,W_N;\mathbb{V}_1,\ldots,\mathbb{V}_\ell) \\
&\le \frac{\ell}{N}\ell!\binom{N}{\ell}\ell F\mathsf{M},
\end{aligned}
$$

where $(a)$ follows from Lemma 8. ■

### C. *Proof of Theorem 4*

For any $\ell \in \{1,\ldots,K\}$, let $\mathcal{Q}_\ell^{\text{rep}}$ be the set of all ordered length-$\ell$ vectors $(d_1,\ldots,d_\ell)\in\mathcal{N}^\ell$, where repetitions are allowed. Notice that:

$$|\mathcal{Q}_\ell^{\text{rep}}| = N^\ell. \tag{25}$$

Recall also that in the average-case scenario under investigation, the demand vector $\mathbf{d} := (d_1,\ldots,d_K)$ is uniform over $\mathcal{Q}_K^{\text{rep}}$. Let $\mathbf{D} := (D_1,\ldots,D_K) \sim$ Uniform($\mathcal{N}^K$).

Fix now an $\ell \in \{1,\ldots,K\}$, and average Inequality (11) over all demand vectors $\mathbf{d}\in\mathcal{Q}_K^{\text{rep}}$. This yields:

$$\mathbb{E}_{\mathbf{D}}[R_{\mathbf{D}}+\epsilon] \ge \mathbb{E}_{\mathbf{D}}[\kappa_{\mathbf{D}}(\ell)] - \sum_{k=1}^{\ell}\beta_k, \tag{26}$$

where

$$\beta_1 := \frac{1}{F} I(W_{D_1};\mathbb{V}_1|\mathbf{D}), \tag{27a}$$

and for $k=2,\ldots,\ell$:

$$\beta_k := \frac{1}{F} I(W_{D_k};\mathbb{V}_1,\ldots,\mathbb{V}_k|W_{D_1},\ldots,W_{D_{k-1}},\mathbf{D}). \tag{27b}$$

The following two lemmas and letting $\epsilon\to 0$ and thus $F\to\infty$, conclude the proof.

*Lemma 10:*

$$\mathbb{E}_{\mathbf{D}}\big[\kappa_{\mathbf{D}}(\ell)\big] = N\Big(1-\Big(1-\frac{1}{N}\Big)^\ell\Big). \tag{28}$$

*Proof:*

$$
\begin{aligned}
\mathbb{E}_{\mathbf{D}}\big[\kappa_{\mathbf{D}}(\ell)\big] &= \mathbb{E}_{\mathbf{D}}\left[\sum_{k=1}^{\ell} \mathbb{1}\big\{D_k \notin \{D_1,\ldots,D_{k-1}\}\big\}\right] \\
&= \sum_{k=1}^{\ell} \mathbb{E}_{\mathbf{D}}\left[\mathbb{1}\big\{D_k \notin \{D_1,\ldots,D_{k-1}\}\big\}\right] \\
&= \sum_{k=1}^{\ell}\left(1-\frac{1}{N}\right)^{k-1} \\
&= N\left(1-\left(1-\frac{1}{N}\right)^\ell\right).
\end{aligned} \tag{29}
$$

■

*Lemma 11:* Parameters $\beta_1, \ldots, \beta_\ell$ satisfy

$$\sum_{k=1}^{\ell} \beta_k \leq \min\left\{\mathbb{E}_{\mathbf{D}}\big[\kappa_{\mathbf{D}}(\ell)\big] \cdot \frac{\ell \mathsf{M}}{N},\ \sum_{k=1}^{\ell} \frac{k\mathsf{M}}{N}\right\}. \tag{30}$$

*Proof:* We first prove that for each $k \in \{1, \ldots, \ell\}$:

$$\beta_k \leq \frac{k\mathsf{M}}{N}, \tag{31}$$

which establishes the upper bound

$$\sum_{k=1}^{\ell} \beta_k \leq \sum_{k=1}^{\ell} \frac{k\mathsf{M}}{N}. \tag{32}$$

Defining $\mathbf{D}_k := (D_1, \ldots, D_k)$, we have:

$$\begin{aligned}
F\beta_k &= I(W_{D_k}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{D_1}, \ldots, W_{D_{k-1}}, \mathbf{D}) \\
&= I(W_{D_k}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{D_1}, \ldots, W_{D_{k-1}}, \mathbf{D}_k) \\
&= \frac{1}{N^k} \sum_{\mathbf{d} \in \mathcal{Q}_k^{\text{rep}}} I(W_{d_k}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{d_1}, \ldots, W_{d_{k-1}}) \\
&= \frac{1}{N^k} \sum_{\tilde{\mathbf{d}} \in \mathcal{Q}_{k-1}^{\text{rep}}} \sum_{j=1}^{N} I(W_j; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{\tilde{\mathbf{d}}}) \\
&\overset{(a)}{\leq} \frac{1}{N^k} \sum_{\tilde{\mathbf{d}} \in \mathcal{Q}_{k-1}^{\text{rep}}} I(W_1, \ldots, W_N; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{\tilde{\mathbf{d}}}) \\
&\overset{(b)}{\leq} \frac{1}{N^k} \sum_{\tilde{\mathbf{d}} \in \mathcal{Q}_{k-1}^{\text{rep}}} kF\mathsf{M} \\
&= \frac{kF\mathsf{M}}{N},
\end{aligned}$$

where $(a)$ holds because the messages are independent and because $H(A_1, \ldots, A_L) \leq \sum_{l=1}^{L} H(A_l)$ for any random $L$-tuple $(A_1, \ldots, A_L)$; and (b) holds because $I(W_1, \ldots, W_N; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{\tilde{\mathbf{d}}})$ cannot exceed $kF\mathsf{M}$. This concludes the proof of (31) and thus (32).

We now prove

$$\sum_{k=1}^{\ell} \beta_k \leq \mathbb{E}_{\mathbf{D}}\big[\kappa_{\mathbf{D}}(\ell)\big] \cdot \frac{\ell \mathsf{M}}{N}. \tag{33}$$

Let $\mathbf{D}_\ell^{\text{dist}}$ be a vector containing all distinct elements of $\mathbf{D}_\ell := (D_1, \ldots, D_\ell)$. Notice that $\mathbf{D}_\ell^{\text{dist}}$ is of length $\kappa_{\mathbf{D}_\ell}(\ell)$. Also, following the definition of the previous section, $W_{\mathbf{D}_\ell} := \{W_{D_1}, \ldots, W_{D_\ell}\} = W_{\mathbf{D}_\ell^{\text{dist}}}$. We have:

$$\begin{aligned}
&F \sum_{k=1}^{\ell} \beta_k \\
&= I(W_{D_1}; \mathbb{V}_1 | \mathbf{D}) \\
&\quad + \sum_{k=2}^{\ell} I(W_{D_k}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{D_1}, W_{D_2}, \ldots, W_{D_{k-1}}, \mathbf{D}) \\
&\leq I(W_{\mathbf{D}_\ell}; \mathbb{V}_1, \ldots, \mathbb{V}_\ell | \mathbf{D}) \\
&\leq I(W_{\mathbf{D}_\ell}; \mathbb{V}_1, \ldots, \mathbb{V}_\ell | \mathbf{D}_\ell) \\
&= I\left(W_{\mathbf{D}_\ell}; \mathbb{V}_1, \ldots, \mathbb{V}_\ell \middle| \mathbf{D}_\ell, \kappa_{\mathbf{D}_\ell}(\ell)\right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{\ell} \mathbb{P}(\kappa_{\mathbf{D}_\ell}(\ell)=i) I\left(W_{\mathbf{D}_\ell}; \mathbb{V}_1,\dots,\mathbb{V}_\ell \middle| \mathbf{D}_\ell, \kappa_{\mathbf{D}_\ell}(\ell)=i\right)\\
&\overset{(a)}{=} \sum_{i=1}^{\ell} \mathbb{P}(\kappa_{\mathbf{D}_\ell}(\ell)=i) I\left(W_{\mathbf{D}_\ell^{\text{dist}}}; \mathbb{V}_1,\dots,\mathbb{V}_\ell \middle| \mathbf{D}_\ell^{\text{dist}}, \kappa_{\mathbf{D}_\ell}(\ell)=i\right)\\
&\overset{(b)}{=} \sum_{i=1}^{\ell} \mathbb{P}(\kappa_{\mathbf{D}_\ell}(\ell)=i) \sum_{\substack{\tilde{\mathbf{d}}\in\mathcal{Q}_i^{\text{dist}}:\\ \tilde{d}_1<\tilde{d}_2\cdots<\tilde{d}_i}} \frac{1}{\binom{N}{i}} I(W_{\tilde{\mathbf{d}}}; \mathbb{V}_1,\dots,\mathbb{V}_\ell)\\
&\overset{(c)}{\leq} \sum_{i=1}^{\ell} \mathbb{P}(\kappa_{\mathbf{D}}(\ell)=i) \frac{i}{N} I(W_1,\dots,W_N; \mathbb{V}_1,\dots,\mathbb{V}_\ell)\\
&\overset{(d)}{\leq} \sum_{i=1}^{\ell} \mathbb{P}(\kappa_{\mathbf{D}}(\ell)=i)\cdot i \cdot \frac{\ell F \mathsf{M}}{N}\\
&= \mathbb{E}_{\mathbf{D}}\left[\kappa_{\mathbf{D}}(\ell)\right]\frac{\ell F\mathsf{M}}{N},
\end{aligned}
$$

where (a) holds because of the Markov chain $\mathbb{V}_1,\dots,\mathbb{V}_\ell - (W_{\mathbf{D}_\ell^{\text{dist}}}, \kappa_{\mathbf{D}_\ell}(\ell) - \mathbf{D}_\ell$; (b) holds because given $\kappa_{\mathbf{D}_\ell}(\ell)=i$ the probability that $\mathbf{D}_\ell^{\text{dist}}$ equals a specific vector $\tilde{\mathbf{d}}\in\mathcal{Q}_i^{\text{dist}}$ equals $\binom{N}{i}^{-1}$; (c) follows from Lemma 8; and (d) follows since $I(W_1,\dots,W_N;\mathbb{V}_1,\dots,\mathbb{V}_\ell)$ cannot be larger than $\ell F\mathsf{M}$. ■

## V. Proof of the Gap-Results in Theorem 2 and Remark 3

### A. *Proof of Theorem 2*

We wish to uniformly bound the gap

$$\xi(K,N,\mathsf{M}) := \frac{R^\star_{\text{worst}}(\mathsf{M})}{R^{\text{low}}_{\text{worst}}(\mathsf{M})}, \tag{34}$$

irrespective of $K,N\geq 1$ and $\mathsf{M}\in[0,N)$.

We recall the achievable rate-memory tradeoff from [14, Corollary 2]. For any pair of positive integers $K,N\geq 1$, define

$$
\begin{aligned}
&R_{\mathsf{YMA}}(K,N,\mathsf{M})\\
&:= \begin{cases} \bar{N} & \text{if } \mathsf{M}=0,\\ \frac{N-\mathsf{M}}{\mathsf{M}}\left(1-\left(1-\frac{\mathsf{M}}{N}\right)^{\bar{N}}\right) & \text{if } \mathsf{M}\in(0,N).\end{cases}
\end{aligned}
$$

Since $R_{\mathsf{YMA}}(K,N,\mathsf{M})$ upper bounds the rate-memory tradeoff under a *decentralized caching* assumption [14], it must also upper bound the rate-memory tradeoff under *centralized caching* as considered here. (In fact, decentralized caching imposes additional constraints on the caching functions $g_k$ compared to our setup here.) Thus, for any number of users $K$ and files $N$:

$$R^\star_{\text{worst}}(\mathsf{M}) \leq R_{\mathsf{YMA}}(K,N,\mathsf{M}), \quad \mathsf{M}\in[0,N). \tag{35}$$

We thus have

$$
\begin{aligned}
\xi(K,N,\mathsf{M}) &\leq \frac{R_{\mathsf{YMA}}(K,N,\mathsf{M})}{R^{\text{low}}_{\text{worst}}(\mathsf{M})}\\
&\leq \frac{R_{\mathsf{YMA}}(K,N,\mathsf{M})}{\underline{R}_{\text{worst}}(K,N,\mathsf{M})},
\end{aligned} \tag{36}
$$

where we defined

$$\underline{R}_{\text{worst}}(K,N,\mathsf{M}) := \max_{\ell\in\bar{\mathcal{N}}} \sum_{j=1}^{\ell}\left(1-\frac{j\mathsf{M}}{N-j+1}\right) \tag{37}$$

and where the second inequality holds because for all $K, N, \mathsf{M}$:

$$\begin{aligned} R^{\text{low}}_{\text{worst}}(\mathsf{M}) &\geq \max_{\ell \in \bar{\mathcal{N}}} \left[ \ell - \sum_{j=1}^{\ell} \frac{j\mathsf{M}}{N-j+1} \right] \\ &= \underline{R}_{\text{worst}}(K, N, \mathsf{M}). \end{aligned} \tag{38}$$

Define

$$\mathsf{M}_\ell := \begin{cases} \frac{N-\ell}{\ell+1} & \text{if } \ell \in \{0, 1, \ldots, \bar{N}-1\}, \\ 0 & \text{if } \ell = \bar{N}. \end{cases}$$

Note that

$$0 = \mathsf{M}_{\bar{N}} < \mathsf{M}_{\bar{N}-1} < \cdots < \mathsf{M}_0 = N.$$

The function $\underline{R}_{\text{worst}}(K, N, \mathsf{M})$ is piecewise-linear with $\bar{N}$ line segments over the intervals

$$[\mathsf{M}_{\ell+1}, \mathsf{M}_\ell], \qquad \ell \in \{1, \ldots, \bar{N}-1\} \tag{39a}$$

$$[\mathsf{M}_1, \mathsf{M}_0), \tag{39b}$$

where the last interval is half-open, see (5).

We next upper bound $R_{\mathsf{YMA}}(K, N, \mathsf{M})$ by a function $\overline{R}_{\mathsf{YMA}}(K, N, \mathsf{M})$ that is piecewise-linear over the same intervals (39). Specifically, for every $\ell \in \bar{\mathcal{N}}$, define for $\mathsf{M} \in [\mathsf{M}_\ell, \mathsf{M}_{\ell-1})$:

$$\begin{aligned} \overline{R}_{\mathsf{YMA}}(K, N, \mathsf{M}) := &\frac{\mathsf{M} - \mathsf{M}_\ell}{\mathsf{M}_{\ell-1} - \mathsf{M}_\ell} \cdot R_{\mathsf{YMA}}(K, N, \mathsf{M}_\ell) \\ &+ \frac{\mathsf{M}_{\ell-1} - \mathsf{M}}{\mathsf{M}_{\ell-1} - \mathsf{M}_\ell} \cdot R_{\mathsf{YMA}}(K, N, \mathsf{M}_{\ell-1}). \end{aligned}$$

Notice that

$$\overline{R}_{\mathsf{YMA}}(K, N, \mathsf{M}_\ell) = R_{\mathsf{YMA}}(K, N, \mathsf{M}_\ell), \quad \forall \ell \in \{1, \ldots, \bar{N}\},$$

whereas for general $\mathsf{M} \in [0, N)$:

$$\overline{R}_{\mathsf{YMA}}(K, N, \mathsf{M}) \geq R_{\mathsf{YMA}}(K, N, \mathsf{M}), \tag{40}$$

because $R_{\mathsf{YMA}}(K, N, \mathsf{M})$ is convex.

Plugging (40) into (36), we obtain:

$$\xi(K, N, \mathsf{M}) \leq \frac{\overline{R}_{\mathsf{YMA}}(K, N, \mathsf{M})}{\underline{R}_{\text{worst}}(K, N, \mathsf{M})} =: \Xi(K, N, \mathsf{M}). \tag{41}$$

Now, since the upper bound $\Xi(K, N, \mathsf{M})$ is continuous and bounded in $\mathsf{M} \in [0, N)$, and because it is quasiconvex [59] in $\mathsf{M}$ over each of the $\bar{N}$ intervals (39),[4] the maximum of $\Xi(K, N, \mathsf{M})$ over each interval is attained at the boundary. Furthermore, since $\overline{R}_{\mathsf{YMA}}(K, N, \mathsf{M}_0) = \underline{R}_{\text{worst}}(K, N, \mathsf{M}_0) = 0$, for $\mathsf{M} \in [\mathsf{M}_1, \mathsf{M}_0)$:

$$\begin{aligned} &\Xi(K, N, \mathsf{M}) \\ &\quad = \frac{\frac{\mathsf{M}-\mathsf{M}_1}{\mathsf{M}_0-\mathsf{M}_1} \overline{R}_{\mathsf{YMA}}(K, N, \mathsf{M}_1) + \frac{\mathsf{M}_0-\mathsf{M}}{\mathsf{M}_0-\mathsf{M}_1} \cdot 0}{\frac{\mathsf{M}-\mathsf{M}_1}{\mathsf{M}_0-\mathsf{M}_1} \underline{R}_{\text{worst}}(K, N, \mathsf{M}_1) + \frac{\mathsf{M}_0-\mathsf{M}}{\mathsf{M}_0-\mathsf{M}_1} \cdot 0} \\ &\quad = \frac{\overline{R}_{\mathsf{YMA}}(K, N, \mathsf{M}_1)}{\underline{R}_{\text{worst}}(K, N, \mathsf{M}_1)} \\ &\quad = \Xi(K, N, \mathsf{M}_1). \end{aligned} \tag{42}$$

Thus, $\Xi(K, N, \mathsf{M})$ is constant over the last interval $[\mathsf{M}_1, \mathsf{M}_0)$ and we obtain that

$$\max_{\mathsf{M} \in [0, N)} \xi(K, N, \mathsf{M}) \leq \max_{\ell \in \bar{\mathcal{N}}} \Xi(K, N, \mathsf{M}_\ell). \tag{43}$$

[4] A linear-fractional function is always quasiconvex.

Irrespective of $K, N \in \mathbb{Z}^+$, we have:

$$\Xi(K, N, \mathsf{M}_{\bar{N}}) = \frac{\bar{N}}{\bar{N}} = 1. \tag{44}$$

When $\bar{N} = 1$ (i.e., only one file or only one user), Inequalities (43) and (44) imply that the gap $\xi(K, N, \mathsf{M}) = 1$ for all $\mathsf{M} \in [0, N)$, and hence our lower bound is exact.

We therefore assume in the following that $\bar{N} \geq 2$. For $\ell \in \{1, \ldots, \bar{N} - 1\}$, we have

$$\begin{aligned}
\Xi(K, N, \mathsf{M}_\ell) &= \frac{\frac{N - \frac{N-\ell}{\ell+1}}{\frac{N-\ell}{\ell+1}} \left(1 - \left(1 - \frac{1}{N}\frac{N-\ell}{\ell+1}\right)^{\bar{N}}\right)}{\sum_{j=1}^{\ell}\left(1 - \frac{j}{N-j+1}\frac{N-\ell}{\ell+1}\right)} \\
&= \frac{\frac{\ell(N+1)}{N-\ell}\left(1 - \left(\frac{\ell(N+1)}{(\ell+1)N}\right)^{\bar{N}}\right)}{\sum_{j=1}^{\ell}\left(1 + \left(1 - \frac{N+1}{N-j+1}\right)\frac{N-\ell}{\ell+1}\right)} \\
&= \frac{\frac{\ell(N+1)}{N-\ell}\left(1 - \left(\frac{\ell(N+1)}{(\ell+1)N}\right)^{\bar{N}}\right)}{\frac{\ell(N+1)}{\ell+1} - \frac{(N-\ell)(N+1)}{\ell+1}\sum_{j=1}^{\ell}\frac{1}{N-j+1}} \\
&= \frac{\frac{\ell(\ell+1)}{N-\ell}\left(1 - \left(\frac{\ell(N+1)}{(\ell+1)N}\right)^{\bar{N}}\right)}{\ell - (N-\ell)\sum_{j=1}^{\ell}\frac{1}{N-j+1}} \\
&\leq \frac{\frac{\ell(\ell+1)}{N-\ell}\left(1 - \left(\frac{\ell(N+1)}{(\ell+1)N}\right)^{N}\right)}{\ell - (N-\ell)\sum_{j=1}^{\ell}\frac{1}{N-j+1}} \\
&\overset{a=\ell/N}{=} \frac{\frac{a(\ell+1)}{(1-a)\ell}\left(1 - \left(\frac{\ell+a}{\ell+1}\right)^{\ell/a}\right)}{1 - (1-a)\sum_{j=0}^{\ell-1}\frac{1}{\ell-aj}} =: \phi(a, \ell).
\end{aligned}$$

Note that since $\ell \in \{1, \ldots, \bar{N} - 1\}$,

$$a \in [1/N, 1). \tag{45}$$

Therefore,

$$\max_{K \in \mathbb{Z}^+} \max_{N \in \mathbb{Z}^+} \max_{\mathsf{M} \in [0,N)} \xi(K, N, \mathsf{M}) \leq \max_{\ell \in \mathbb{Z}^+} \max_{a \in (0,1)} \phi(a, \ell),$$

which concludes the proof.

## B. *Proof of Inequality* (7)

We have a closer look at the denominator of the function $\phi(a, \ell)$.

Notice that $\frac{1}{n} \leq \int_{n-1}^{n} \frac{dt}{t}$ for all $n \geq 2$. Therefore,

$$\begin{aligned}
\sum_{j=0}^{\ell-1}\frac{1}{\ell - aj} &= \sum_{j=0}^{\ell-1}\frac{1}{\ell(1-a) + a(\ell - j)} \\
&\overset{i=\ell-j}{=} \sum_{i=1}^{\ell}\frac{1}{\ell(1-a) + ai} \\
&= \frac{1}{\ell(1-a) + a} + \frac{1}{a}\sum_{i=2}^{\ell}\frac{1}{\ell(1-a)/a + i}
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{\ell(1-a)+a} + \frac{1}{a}\sum_{i=2}^{\ell}\int_{\ell(1-a)/a+i-1}^{\ell(1-a)/a+i} \frac{1}{t}\,dt \\
&= \frac{1}{\ell(1-a)+a} + \frac{1}{a}\int_{\ell(1-a)/a+1}^{\ell(1-a)/a+\ell} \frac{1}{t}\,dt \\
&= \frac{1}{\ell(1-a)+a} + \frac{1}{a}\ln\left(\frac{\ell}{\ell(1-a)+a}\right).
\end{aligned} \tag{46}
$$

We use (46) to upper bound the function $\phi(a,\ell)$:

$$
\begin{aligned}
\phi(a,\ell) &= \frac{\frac{a(\ell+1)}{(1-a)\ell}\left(1-\left(\frac{\ell+a}{\ell+1}\right)^{\ell/a}\right)}{1-(1-a)\sum_{j=0}^{\ell-1}\frac{1}{\ell-aj}} \\
&\leq \frac{\frac{a(\ell+1)}{(1-a)\ell}\left(1-\left(\frac{\ell+a}{\ell+1}\right)^{\ell/a}\right)}{1-\frac{1-a}{\ell(1-a)+a}-\frac{1-a}{a}\ln\left(\frac{\ell}{\ell(1-a)+a}\right)} \\
&\stackrel{b=1/\ell}{=} \frac{\frac{a(1+b)}{1-a}\left(1-\left(\frac{1+ab}{1+b}\right)^{\frac{1}{ab}}\right)}{1-\frac{(1-a)y}{1-a+ab}+\frac{1-a}{a}\ln\left(1-a+ab\right)} = \psi(a,b).
\end{aligned}
$$

Noting also that if $\ell > 10^4$, then $b < 10^{-4}$, this concludes the proof of (7).

## VI. Proof of the Gap-Result in Theorem 5

We wish to uniformly bound the gap

$$
\theta(K,N,\mathsf{M}) := \frac{R^{\star}_{\text{avg}}(\mathsf{M})}{R^{\text{low}}_{\text{avg}}(\mathsf{M})}, \tag{47}
$$

irrespective of $K,N \geq 1$ and $\mathsf{M} \in [0,N)$.

Since $R_{\mathsf{YMA}}(K,N,\mathsf{M})$ upper bounds the rate-memory tradeoff for the worst case, it must also upper bound the rate-memory tradeoff for the average case. Thus, for any number of users $K$ and files $N$:

$$
R^{\star}_{\text{avg}}(\mathsf{M}) \leq R_{\mathsf{YMA}}(K,N,\mathsf{M}), \quad \mathsf{M} \in [0,N). \tag{48}
$$

We thus have

$$
\begin{aligned}
\theta(K,N,\mathsf{M}) &\leq \frac{R_{\mathsf{YMA}}(K,N,\mathsf{M})}{R^{\text{low}}_{\text{avg}}(\mathsf{M})} \\
&\leq \frac{R_{\mathsf{YMA}}(K,N,\mathsf{M})}{\underline{R}_{\text{avg}}(K,N,\mathsf{M})},
\end{aligned} \tag{49}
$$

where we defined

$$
\underline{R}_{\text{avg}}(K,N,\mathsf{M}) := \max_{\ell\in\bar{\mathcal{N}}}\sum_{k=1}^{\ell}\left[\left(1-\frac{1}{N}\right)^{k-1} - \frac{k}{N}\mathsf{M}\right],
$$

and where the second inequality holds because for all $K,N,\mathsf{M}$:

$$
\begin{aligned}
R^{\text{low}}_{\text{avg}}(\mathsf{M}) &\geq \max_{\ell\in\bar{\mathcal{N}}}\left[\left(1-\left(1-\frac{1}{N}\right)^{\ell}\right)N - \frac{\ell(\ell+1)}{2N}\mathsf{M}\right] \\
&= \max_{\ell\in\bar{\mathcal{N}}}\sum_{k=1}^{\ell}\left[\left(1-\frac{1}{N}\right)^{k-1} - \frac{k}{N}\mathsf{M}\right].
\end{aligned}
$$

Define

$$\tilde{\mathsf{M}}_\ell := \begin{cases} \frac{N}{\ell+1}\left(1-\frac{1}{N}\right)^\ell & \text{if } \ell \in \{0,1,\dots,\bar{N}-1\}, \\ 0 & \text{if } \ell = \bar{N}. \end{cases}$$

Note that $0 = \tilde{\mathsf{M}}_{\bar{N}} < \tilde{\mathsf{M}}_{\bar{N}-1} < \cdots < \tilde{\mathsf{M}}_0 = N$. The function $\underline{R}_{\text{avg}}(K,N,\mathsf{M})$ is piecewise-linear with $\bar{N}$ line segments over the intervals

$$[\tilde{\mathsf{M}}_{\ell+1},\ \tilde{\mathsf{M}}_\ell], \qquad \ell \in \{1,\dots,\bar{N}-1\} \tag{50a}$$

$$[\tilde{\mathsf{M}}_1,\ \tilde{\mathsf{M}}_0), \tag{50b}$$

where the last interval is half-open, see (5).

We next upper bound $R_{\mathsf{YMA}}(K,N,\mathsf{M})$ by a function $\overline{R}_{\mathsf{YMA}}(K,N,\mathsf{M})$ that is piecewise-linear over the same intervals (50). Specifically, for every $\ell \in \bar{\mathcal{N}}$, define for $\mathsf{M} \in [\tilde{\mathsf{M}}_\ell, \tilde{\mathsf{M}}_{\ell-1})$:

$$\begin{aligned}\overline{R}_{\mathsf{YMA}}(K,N,\mathsf{M}) := & \frac{\mathsf{M}-\tilde{\mathsf{M}}_\ell}{\tilde{\mathsf{M}}\ell-1-\tilde{\mathsf{M}}_\ell}\cdot R_{\mathsf{YMA}}(K,N,\tilde{\mathsf{M}}_\ell) \\ & + \frac{\tilde{\mathsf{M}}_{\ell-1}-\mathsf{M}}{\tilde{\mathsf{M}}_{\ell-1}-\tilde{\mathsf{M}}_\ell}\cdot R_{\mathsf{YMA}}(K,N,\tilde{\mathsf{M}}_{\ell-1}).\end{aligned}$$

Notice that

$$\overline{R}_{\mathsf{YMA}}(K,N,\tilde{\mathsf{M}}_\ell) = R_{\mathsf{YMA}}(K,N,\tilde{\mathsf{M}}_\ell), \quad \forall \ell \in \{1,\dots,\bar{N}\},$$

whereas for general $\mathsf{M} \in [0,N)$:

$$\overline{R}_{\mathsf{YMA}}(K,N,\mathsf{M}) \geq R_{\mathsf{YMA}}(K,N,\mathsf{M}), \tag{51}$$

because $R_{\mathsf{YMA}}(K,N,\mathsf{M})$ is convex.

Plugging (51) into (49), we obtain:

$$\theta(K,N,\mathsf{M}) \leq \frac{\overline{R}_{\mathsf{YMA}}(K,N,\mathsf{M})}{\underline{R}_{\text{avg}}(K,N,\mathsf{M})} =: \Theta(K,N,\mathsf{M}). \tag{52}$$

Following similar arguments as in the proof of Theorem 2, we have

$$\max_{\mathsf{M}\in[0,N)} \theta(K,N,\mathsf{M}) \leq \max_{\ell\in\bar{\mathcal{N}}} \Theta(K,N,\tilde{\mathsf{M}}_\ell). \tag{53}$$

Irrespective of $K,N \in \mathbb{Z}^+$, we have:

$$\begin{aligned}\Theta(K,N,\tilde{\mathsf{M}}_{\bar{N}}) &= \frac{\bar{N}}{N(1-(1-1/N)^{\bar{N}})} \\ &= \frac{x}{1-((1-1/N)^N)^x}\Big|_{x=\bar{N}/N} \\ &\overset{(a)}{\leq} \frac{x}{1-e^{-x}}\Big|_{x=\bar{N}/N} \\ &\overset{(b)}{\leq} \frac{1}{1-e^{-1}},\end{aligned}$$

where $(a)$ follows since $(1-1/\zeta)^\zeta \leq e^{-1}$ for all $\zeta > 1$ and $(b)$ follows since $x \mapsto \frac{x}{1-e^{-x}}$ is an increasing function. This implies that when $\bar{N} = 1$ (i.e., one file or one user), then $\Theta(K,N,\mathsf{M}) \leq \frac{1}{1-e^{-1}} \leq 1.582$ for all $\mathsf{M} \in [0,N)$.

In the following, we assume that $\bar{N} \geq 2$. As for $\ell \in \{1,\dots,\bar{N}-1\}$, we have

$$\begin{aligned}&\Theta(K,N,\mathsf{M}_\ell) \\ &= \frac{\frac{N-\frac{N}{\ell+1}\left(1-\frac{1}{N}\right)^\ell}{\frac{N}{\ell+1}\left(1-\frac{1}{N}\right)^\ell}\left(1-\left(1-\frac{1}{N}\frac{N}{\ell+1}\left(1-\frac{1}{N}\right)^\ell\right)^{\bar{N}}\right)}{N-\left(N+\frac{\ell}{2}\right)\left(1-\frac{1}{N}\right)^\ell}\end{aligned}$$

$$
\begin{aligned}
&= \frac{\left(\frac{\ell+1}{N} - \frac{1}{N}\left(1-\frac{1}{N}\right)^{\ell}\right)\left(1-\left(1-\frac{1}{\ell+1}\left(1-\frac{1}{N}\right)^{\ell}\right)^{\bar{N}}\right)}{\left(1-\frac{1}{N}\right)^{\ell}\left(1-\left(1+\frac{\ell}{2N}\right)\left(1-\frac{1}{N}\right)^{\ell}\right)} \\
&\le \frac{\left(\frac{\ell+1}{N} - \frac{1}{N}\left(1-\frac{1}{N}\right)^{\ell}\right)\left(1-\left(1-\frac{1}{\ell+1}\left(1-\frac{1}{N}\right)^{\ell}\right)^{N}\right)}{\left(1-\frac{1}{N}\right)^{\ell}\left(1-\left(1+\frac{\ell}{2N}\right)\left(1-\frac{1}{N}\right)^{\ell}\right)} \\
&\overset{(a)}{=} \frac{\left(u+v-v\left(1-v\right)^{u/v}\right)\left(1-\left(1-\frac{v}{u+v}\left(1-v\right)^{u/v}\right)^{1/v}\right)}{\left(1-v\right)^{u/v}\left(1-\left(1+\frac{u}{2}\right)\left(1-v\right)^{u/v}\right)} \\
&=: \eta(u,v),
\end{aligned}
$$

where $(a)$ follows by a change of variable $u = \ell/N$ and $v = 1/N$. Note that since $\ell \in \{1, \ldots, \bar{N}-1\}$ and assuming $\bar{N} \ge 2$,

$$u \in \left[\frac{1}{N}, \frac{\bar{N}-1}{N}\right] \text{ and } v \in (0, 1/2].$$

Also, it holds that $\frac{1}{1-e^{-1}} < \max_{u\in(0,1]} \max_{v\in(0,1/2]} \eta(u,v)$. Therefore,

$$\max_{K\in\mathbb{Z}^+} \max_{N\in\mathbb{Z}^+} \max_{\mathsf{M}\in[0,N)} \theta(K,N,\mathsf{M}) \le \max_{u\in(0,1]} \max_{v\in(0,1/2]} \eta(u,v),$$

which concludes the proof.

## References


[1] M. A. Maddah-Ali, U. Niesen, "Fundamental limits of caching," in *IEEE Trans. Inform. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.
[2] Z. Chen, P. Fan, and K. B. Letaief, "Fundamental limits of caching: Improved bounds for small buffer users." Online: `ArXiv:1407.1935`.
[3] C. Tian, "A note on the fundamental limits of coded caching," *arXiv*, 1503.00010v1, Feb. 2015.
[4] K. Wan, D. Tuninetti, and P. Piantanida, "On the Optimality of Uncoded Cache Placement," *arXiv*,1511.02256, November, 2015.
[5] K. Wan, D. Tuninetti, and P. Piantanida, "On caching with more users than files," *arXiv*, 1601.063834v2, Jan., 2016.
[6] A. Sengupta, R. Tandon, and T. C. Clancy, "Improved approximation of storage-rate tradeoff for caching via new outer bounds," in Proc. of *IEEE ISIT*, Hong Kong, June 2015.
[7] S. Sahraei and M. Gastpar, "K users caching two files: An improved achievable rate," in Proc. *CISS*, pp. 620–624, Mar. 2016.
[8] C. Tian and J. Chen, "Caching and delivery via interference elimination," Online: *arXiv*, 1604.08600, 2016.
[9] H. Ghasemi and A. Ramamoorthy, "Improved lower bounds for coded caching," Online: `Arxiv:1501.06003`.
[10] H. Ghasemi and A. Ramamoorthy, "Further results on lower bounds for coded caching," in Proc. *ISIT 2016*.
[11] C. Tian, "Symmetry, demand types and outer bounds in caching systems," in Proc. *IEEE ISIT*, 2016.
[12] A. Amiri, Q. Yang, and D. Gündüz, "Coded caching for a large number of users," in Proc. *ITW 2016*.
[13] C.-Y. Wang, S. H. Lim, and M. Gastpar, "A new converse bound for coded caching," in *Proc. of ITA 2016*, La Jolla (CA), Feb. 2016.
[14] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "The exact rate-memory tradeoff for caching with uncoded prefetching," in *arXiv:1609.07817 [cs.IT]*, Sep. 2016.
[15] J. Gómez-Vilardebó, "Fundamental limits of caching: improved bounds with coded prefetching," in *arXiv:1612.09071v2[cs.IT]*, Jan. 2017.
[16] M. M. Amiri, Q. Yang, D. Gündüz, "Decentralized Coded Caching with Distinct Cache Capacities," *arXiv*, 1610.03792, Oct., 2016.
[17] S. Wang, W. Li, X. Tian, and H. Liu, "Fundamental limits of heterogenous cache," *arXiv*, 1504.01123v1, Apr., 2015.
[18] R. Pedarsani, M. A. Maddah-Ali and U. Niesen, "Online coded caching," *to appear IEEE/ACM Trans. Networking*.
[19] U. Niesen and M. A. Maddah-Ali, "Coded caching with nonuniform demands," in *Proc. IEEE INFOCOM Workshop*, 2014.
[20] M. Ji, A. M. Tulino, J. Llorca, and G. Caire, "Order-optimal rate of caching and coded multicasting with random demands." Online: `ArXiv:11502.03124`.
[21] J. Zhang, X. Lin, and X. Wang, "Coded caching under arbitrary popularity distributions," in Proc. *ITA*, pp. 98–107, Feb. 2015
[22] R. Timo, S. Saeedi Bidokhti, M. Wigger and B. Gieger, "A rate-distortion approach to caching," in proceedings *International Zurich Seminar on Communications (IZS)*, Zurich, Switzerland, March, 2016.
[23] C.-Y. Wang, S. H. Lim, and M. Gastpar, "Information-theoretic caching: Sequential coding for computing," *arXiv*, 1504.00553, April, 2015.
[24] S. H. Lim, C.-Y. Wang, and M. Gastpar, "Information theoretic caching: The multi-user case," *arXiv* 1604.02333, 2016.
[25] R. Timo and M. Wigger, "Joint cache-channel coding over erasure broadcast channels," in *Proc. ISWCS 2015*, Bruxelles, Belgium, Aug. 2015.

[26] S. Saeedi Bidokhti, M. Wigger, and R. Timo, “Noisy Broadcast Networks with Receiver Caching,” submitted to *IEEE Trans. Inf. Theory*, *arXiv*, 1605.02317, May, 2016.
[27] S. Saeedi Bidokhti, M. Wigger, and R. Timo, “An Upper Bound on the Capacity-Memory Tradeoff of Degraded Broadcast Channels,” in Proc. of *IEEE ISTC*, Brest, Sep., 2016.
[28] S. Saeedi Bidokhti, M. Wigger, and A. Yener, “Benefits of cache-assignment on noisy broadcast channels,” submitted to *ISIT'2017*.
[29] W. Huang, S. Wang, N. Ding, F. Yang, and W. Zhang, “The performance analysis of coded cache in wireless fading channel.” Online: `ArXiv:1504.01452`.
[30] S. Wang, X. Tian and H. Liu, “Exploiting the unexploited of coded caching for wireless content distribution,” in *Proc. IEEE ICNC, 2015.*
[31] P. Hassanzadeh, E. Erkip, J. Llorca,and A. Tulino, “Distortion-memory tradeoffs in Cache-aided wireless video delivery,” in *Proc. of Allerton Conference* Monticello (IL), USA, Oct. 2015.
[32] A. S. Cacciapuoti, M. Caleffi, M. Ji, J. Llorca, A. M. Tulino,“Speeding up future video distribution via channel-aware caching-aided coded multicast,” *IEEE JSAC in Comm.*, vol. 34, no. 8, Aug., 2016.
[33] A. Ghorbel, M. Kobayashi, and S. Yang, “Cache-enabled broadcast packet erasure channels with state feedback,” in *Proc. of Allerton Conference,* Monticello (IL), USA, Oct. 2015.
[34] A. Ghorbel, M. Kobayashi, and S. Yang “Content delivery in erasure broadcast channels with cache and feedback,” *IEEE Trans. on Inf. Theory*, vol. 62, no. 11, pp. 6407–6422, Nov. 2016.
[35] A. Ghorbel, M. Kobayashi, and S. Yang, “Content delivery in erasure broadcast channels with cache and feedback,” in Proc. *ISIT 2016*, pp. 835–839, July 2016.
[36] S. Yang, K.-H. Ngo, and M. Kobayashi, “Content delivery with coded caching and massive MIMO in 5G,” in Proc. of *9th ISTC*, Sep. 2016, pp. 370–374.
[37] J. Zhang and P. Elia, “Fundamental limits of cache-aided wireless BC: interplay of coded-caching and CSIT feedback.” Online: `ArXiv:1511.03961`.
[38] J. Hachem, N. Karamchandani, and S. Diggavi, “Content caching and delivery over heterogeneous wireless networks,” *arXiv*, 1404.6560, March, 2015.
[39] J. Zhang, F. Engelmann, and P. Elia, “Coded caching for reducing CSIT-feedback in wireless communications,” in Proc. *Allerton Conf.*, pp. 1099–1105, 2015.
[40] J. Zhang and P. Elia, “Wireless coded caching: A topological perspective,” *arXiv* 1606.08253.
[41] J. Zhang and P. Elia, “Feedback-aided coded caching for the MISO BC with small caches,” *arXiv* 1606.05396.
[42] X. Yi and G. Caire, “Topological coded caching,” in Proc. *ISIT*, July 2016, pp. 2039–2043.
[43] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, “Edge-facilitated wireless distributed computing,” in Proc. *IEEE GLOBECOM*, Dec. 2016.
[44] M. A. Maddah-Ali and U. Niesen, “Cache-aided interference channels,” in *Proc. IEEE Intl. Symp. Inform. Theory,* Hong Kong China, June 2015.
[45] N. Naderializadeh, M. A. Maddah-Ali, and A. S. Avestimehr, “Fundamental limits of cache-aided interference management,” *arXiv*, 1602.04207, February 2016.
[46] S. Pooya Shariatpanah, S. Abolfazl Motahari, B. Hossein Khalaj, “Multi-server coded caching,” *IEEE Trans. on Inf. Theory*, vol. 62, no. 12, pp. 7253–7271, Dec. 2016.
[47] M. Wigger, R. Timo, and S. Shamai, “Complete Interference Mitigation through Receiver-Caching in Wyner's Networks,” in *Inf. Theory Workshop*, Cambridge, Sep., 2016.
[48] V. Ntranos, M. A. Maddah-Ali, and G. Caire, “Cellular interference alignment,” *IEEE Trans. Inform. Theory,* vol. 61, no. 3, March 2015, pp. 1194–1217.
[49] Y. Ugur, Z. H. Awan and A. Sezgin, “Cloud radio access networks with coded caching.” Online: `ArXiv:1512.02385`.
[50] S.-H. Park, O. Simeone, and S. Shamai (Shitz), “Joint optimization of cloud and edge processing for fog radio access networks.” Online: `ArXiv:1601.02460`
[51] R. Tandon and O. Simeone, “Cloud-aided wireless networks with edge caching: Fundamental latency trade-offs in fog radio access networks.” Jan. 2016.
[52] B. Azari, O. Simeone, U. Spagnolini and A. Tulino, ”Hypergraph-based analysis of clustered cooperative beamforming with application to edge caching.” to appear in *IEEE Wireless Comm. Letters*.
[53] M. Peng, S. Yan, K. Zhang and C. Wang, “Fog computing based radio access networks: issues and challenges.” Online: `ArXiv:1506.04233`.
[54] N. Karamchandani, U. Niesen, M. A. Maddah-Ali, and S. Diggavi, “Hierarchical coded caching,” in Proc. *IEEE ISIT*, pp. 2142–2146, June 2014.
[55] J. Hachem, N. Karamchandani, and S. Diggavi, “Effect of number of users in multi-level coded caching,” in Proc. *IEEE ISIT*, June 2015.
[56] J. Hachem, N. Karamchandani and S. Diggavi, “Multi-level coded caching,” in proceedings *IEEE ISIT*, Honolulu (HI), USA, July, 2014.
[57] S. Sahraei and M. Gastpar, “Multi-library coded caching,” *arXiv*, 1601.06016, January, 2016.
[58] B. Blaszczyszyn and A. Giovanidis, “Optimal geographic caching in cellular networks,” *arXiv*, 1409.7626, September, 2014.
[59] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge Univ. Press, 2004.
[60] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2006.